

## PENGENALAN CITRA SIDIK JARI MENGGUNAKAN METODE *PRINCIPAL COMPONENT ANALYSIS* DAN *HAMMING DISTANCE*

Samuel Lukas, Meiliyana, Gunawan Sugianto  
Fakultas Ilmu Komputer, Universitas Pelita Harapan, Indonesia  
UPH Tower, Lippo Karawaci, Tangerang, Indonesia  
e-mail: slukas@uph.edu, meiliyana@uph.edu

### ABSTRACT

Each person has a unique fingerprint and many research have been conducted to recognize these fingerprints. Many applications have also been implemented related to fingerprint matching. This paper discusses on how good the *Principal Component Analysis (PCA)* which is combined with *Hamming distance* in recognizing five fingerprints.

Some experiments are conducted to get what is the best number of *PCA* vector in recognizing the five human fingerprints data. The result of these experiments are also reported in this paperinally, the results of all experiments are also be reported.

**Keywords:** *Principal Component Analysis, finger print, Hamming distance.*

### 1. INTRODUCTION

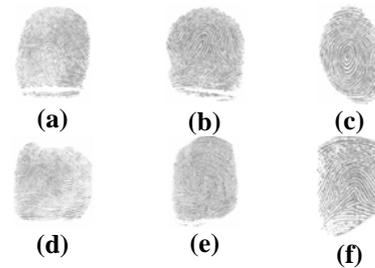
Sistem keamanan saat ini banyak menggunakan sidik jari sebagai salah satu media akses. Sistem keamanan menggunakan sidik jari sangat efisien dibandingkan dengan sistem keamanan konvensional lainnya seperti *password*, kode batang, kartu pntar dan lainnya. Akan tetapi penerapan sistem pegidentifikasi citra sidik ada juga keterbatasannya misalnya masalah seperti *noise* pada citra sehingga citra sidik jari menjadi kurang jelas.

Untuk menyederhanakan penelitian maka pengumpulan sidik jari tidak digunakan alat scanner langsung ke sidik jari akan tetapi sidik jari seseorang didata melalui selembar kertas dan kemudian hasilnya diubah kedalam bentuk elektronik untuk diproses lebih lanjut. Bentuk elektronik citra sidik jari disimpan dalam bentuk *bmp* dengan ukuran 150 x 100 pixel berderajat keabuan 8 bit. Untuk setiap orang disimpan 10 data elektronik citra sidik ibu jari tangan kanannya. Sistem pengenalan dikembangkan dengan bahasa *Basic 6.0* dan piranti lunak aplikasi *Matlab 6.5* untuk mencari vektor eigen dari matrik data.

### 2. LANDASAN TEORI

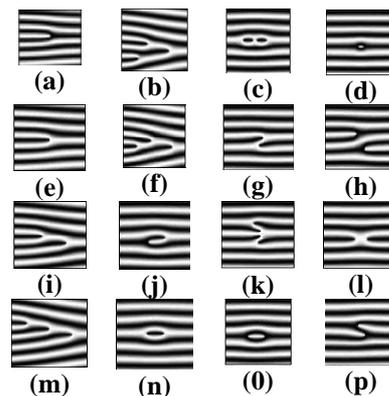
Pengklasifikasian sidik jari manusia pola garis atau pola titik yang disebut dengan *minuteae*. Pola alur pada sidik jari terbagi menjadi beberapa jenis yaitu *loop*, *whorl*, *arch* yang diperlihatkan pada Gambar 1, sedangkan pola titik pada sidik jari dibagi menjadi beberapa tipe yang diperlihatkan pada Gambar 2.

*Principal Component Analysis (PCA)* merupakan teknik untuk merepresentasikan sebuah objek, dengan cara mengekstraksi ciri-ciri sebuah objek dan mereduksi dimensi dari objek tersebut dengan cara mentransformasikan sejumlah variabel korelasi ke jumlah yang lebih sedikit [2] yang dikenal dengan *feature extraction*.



**Gambar 1.** Pola alur sidik jari.

(a). *Left-leaning Loop* (b). *Right-leaning Loop*  
(c). *Whorl* (d). *Double Loop* (e). *Double Loop with central pocket* (f). *Tented Arch*



**Gambar 2.** Pola titik sidik jari.

(a). *Beginning or Ending* (b). *Triple bifurcation II*  
(c). *Double whorl* (d). *Point* (e). *Single bifurcation*  
(f). *Triple bifurcation III* (g). *Single bridge*  
(h). *Through line* (i). *Double bifurcation*  
(j). *Hook* (k). *Twin bridge* (l). *Crossing*  
(m). *Triple bifurcation I* (n). *Single whorl*  
(o). *Interval* (p). *Side contact*

Diberikan  $M$  buah citra, yang direpresentasikan kedalam bentuk matrik baris

berukuran  $I \times N$  dan kemudian dibentuk matrik data  $X$ .

$$X_i = (x_{i,1} \ x_{i,2} \ x_{i,3} \ \dots \ x_{i,N-1} \ x_{i,N}) \quad i = 1,2,3,\dots,M$$

$$X^t = (X_1 \ X_2 \ X_3 \ \dots \ X_{M-1} \ X_M)$$

Kemudian dibentuk matrik baris  $\Psi$  berordo  $I \times N$  yang merupakan rata-rata kolom dari matrik  $X$  sehingga

$$\Psi = (\varphi_1, \varphi_2, \varphi_3, \dots, \varphi_N)$$

$$\varphi_j = \frac{1}{M} \sum_{i=1}^M x_{i,j} \quad j = 1,2,3,\dots,N$$

Selanjutnya adalah membentuk matriks *covariance*  $\Phi$  dimana  $\Phi$  merupakan selisih matriks data,  $X$ , dengan matriks  $\Psi$ .

$$\Phi^t = (\Phi_1, \Phi_2, \Phi_3, \dots, \Phi_M)$$

$$\Phi_i = X_i - \Psi \quad i = 1,2,3,\dots,M$$

$$Covariance = \frac{1}{M} \sum_{n=1}^M \Phi_n \Phi_n^t$$

Dari matrik *covariance* yang berordo  $M \times M$  ini, dicari nilai *eigen* (*eigenvalue*) dan nilai vektor *eigen* (*eigenvector*) dengan menggunakan program aplikasi *Matlab*.

*Hamming Distance* adalah salah satu metoda pengujian untuk mencari seberapa mirip sebuah vektor terhadap vektor lainnya berdasarkan nilai kedekatannya. Jika nilai kedekatan semakin kecil maka artinya kemiripan kedua vektor semakin besar sebaliknya jika nilai kedekatan semakin besar artinya kemiripan kedua vektor semakin kecil.

### 3. IMPLEMENTASI DAN PENGUJIAN

Implementasi dari pengenalan citra sidik jari menggunakan metoda *principle component analysis* dan *hamming distance* dilakukan dengan tiga tahapan. Tahapan pertama membentuk matrik data referensi, Norm berordo  $M \times M$ . Tahapan kedua membentuk matrik Data yang akan diujikan berordo  $I \times M$ . Tahapan ketiga adalah pengujian. Langkah pada setiap tahapan diperlihatkan sebagai berikut

Tahapan pembentukan matrik data referensi, Norm.

1. Pembentukan matrik data citra sidik jari
2. Pembentukan matrik *covariance* citra sidik jari
3. Perhitungan nilai *eigen* dan vektor *eigen* dari matrik *covariance*.
4. Pengurutan vektor *eigen* dari besar ke kecil berdasarkan nilai *eigen*. ( $V_{sort}$ )
5. Bentuk matrik  $Construct = \Phi^t x V_{Sort}$  yang berordo  $N \times M$
6. Bentuk marik  $E$ ,  $Extract = \Phi x Construct$  yang berordo  $M \times M$
7. Bentuk matrik Normalisasi, Norm, dari matrik  $Extract$

$$Norm = \begin{bmatrix} N_{1,1} & N_{1,2} & N_{1,3} & \dots & N_{1,50} \\ N_{2,1} & N_{2,2} & N_{2,3} & \dots & N_{2,50} \\ N_{3,1} & N_{3,2} & N_{3,3} & \dots & N_{3,50} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ N_{50,1} & N_{50,2} & N_{50,3} & \dots & N_{50,50} \end{bmatrix}, \quad N_{i,j} = \begin{cases} +1 & E_{i,j} \geq 0 \\ -1 & \text{lainnya} \end{cases}$$

Tahapan pembentukan matrik Data.

1. Pembentukan matrik data  $Y$ , berordo  $I \times N$ .
2. Pembentukan matrik  $Temp = Y - \Psi$  berordo  $I \times N$ .
3. Pembentukan matrik  $Test = Temp \times Construct$  berordo  $I \times M$ .
4. Bentuk matrik Data dengan melakukan normalisasi data.

$$Data = [d_1 \ d_2 \ d_3 \ \dots \ d_M] \quad d_i = \begin{cases} +1 & test_i \geq 0 \\ -1 & \text{lainnya} \end{cases}$$

Tahapan pengujian data dilakukan dengan membandingkan matrik Data dengan Matrik Norm dengan langkah berikut ini. .... (5)

1. Tentukan nilai  $h_i = \frac{1}{2} \sum_{j=1}^M |d_j + N_{i,j}| \quad i = 1,2,3,\dots,M$
2. Data adalah objek ke  $i$  dari data referensi apabila  $h_i$  adalah terbesar.

Rancangan antarmuka sistem pengenalan ini dimulai dari form utama yang diperlihatkan pada Gambar 3. Tombol *set data reference* merepresentasikan tahapan pertama. Tombol *Testing* adalah representasi pada tahapan kedua dan ketiga. Tombol *Experimental* melakukan rangkuman atas beberapa percobaan yang dilakukan.

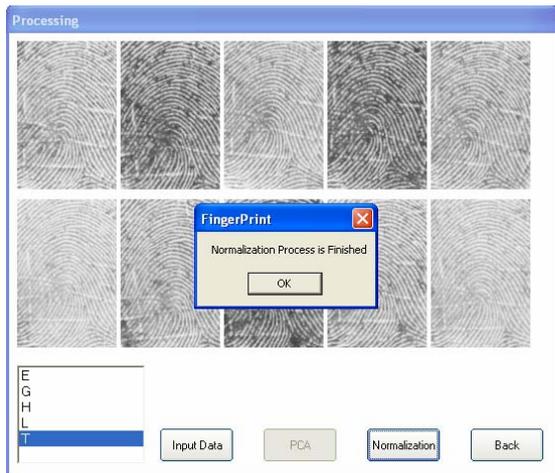


Gambar 3. form utama.

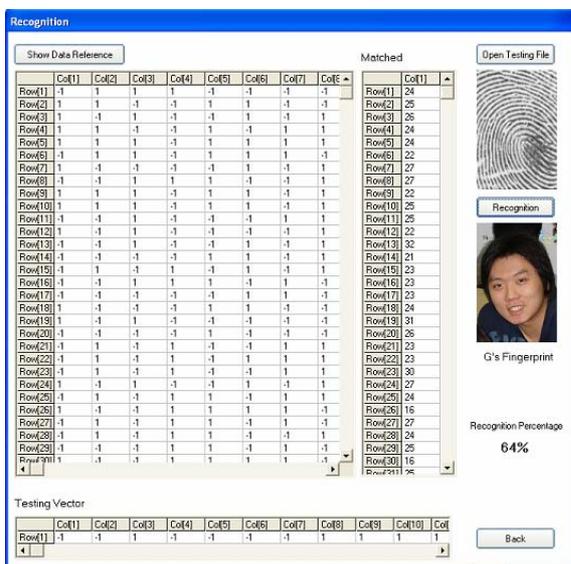
Apabila tombol *set data reference* ditekan maka akan ditampilkan menu *form processing* yang diperlihatkan pada Gambar 4.

Tombol input data dilakukan untuk memilih data referensi yang akan diproses yang kemudian dilakukan pemrosesan dengan PCA dan diakhiri dengan tombol normalisasi untuk menormalisasi data referensi

Apabila tombol *Testing* pada form utama ditekan maka akan tampil form recognition yang diperlihatkan pada Gambar 5. Tombol *Show Data Reference* menampilkan hasil normalisasi data pada proses sebelumnya. Tombol *open testing file* digunakan untuk memilih citra sidik jari yang ingin dikenali. Setelah memilih citra sidik jari yang ingin dikenali, maka proses identifikasi bisa dilakukan dengan memilih tombol *recognition*.



Gambar 4. form processing



Gambar 5. form recognition.

Beberapa pengujian dilakukan untuk menentukan berapa banyak jumlah eigen vektor minimum untuk suatu vektor data dengan ketelitian yang masih dapat diterima oleh sistem. Untuk itu maka dilakukan tiga pengujian dengan mengambil 50, 40 dan 30 data eigen vektor untuk setiap gambar sidik jari. Citra sidik jari yang akan diuji diambil dari data yang tidak ada dalam data sidik jari pada referensi.

Tabel hasil pengujian untuk 50, 40 dan 30 data eigen vektor diperlihatkan pada Tabel 1, Tabel 2 dan Tabel 3.

Tabel 1. Citra sidik jari yang dikenal [50 x 50]

	E	G	H	L	T
E	6	0	0	0	4
G	4	4	0	1	1
H	0	0	10	0	0
L	0	3	2	4	1
T	0	0	1	0	9

Tabel 2. Citra sidik jari yang dikenal [50 x 40]

	E	G	H	L	T
E	8	0	1	0	1
G	1	6	2	1	0
H	0	0	9	0	1
L	0	2	0	7	1
T	0	0	0	0	10

Tabel 3. Citra sidik jari yang dikenal [50 x 30]

	E	G	H	L	T
E	8	0	1	0	1
G	1	6	0	3	0
H	0	0	8	1	1
L	0	3	0	6	1
T	1	0	0	0	9

Setelah melakukan ketiga pengujian dengan data referensi yang berbeda, dapat diambil kesimpulan bahwa data referensi dengan ukuran 50x40 memiliki tingkat akurasi yang paling tinggi, dimana persentase keakuratan dapat dilihat pada tabel 4.

Tabel 4. Banyak citra sidik jari yang dikenal

	50x30 Data referensi	50x40 Data referensi	50x50 Data referensi
E	8	8	6
G	6	6	4
H	8	9	10
L	6	7	4
T	9	10	9
<b>Persentase</b>	<b>74%</b>	<b>80%</b>	<b>66%</b>

#### 4. KESIMPULAN DAN SARAN

Berdasarkan proses pengujian yang telah dilakukan, dapat diambil kesimpulan sebagai berikut:

1. PCA sangat baik digunakan untuk mereduksi dimensi citra sehingga mengurangi kompleksitas citra tetapi tidak mengurangi informasi penting yang ada pada setiap citra.
2. Dengan menggunakan teknik PCA dan metode *Hamming Distance*, sidik jari telah dapat dikenali dengan baik, walaupun sidik jari tersebut belum terdapat pada *database*.
3. Berdasarkan penelitian yang telah dilakukan, dapat diketahui bahwa pengenalan sidik jari dengan matriks data referensi berukuran 50x40 menghasilkan tingkat pengenalan yang paling tinggi yaitu dengan persentase pengenalan mencapai 80%.

Beberapa saran yang dapat diajukan untuk mengembangkan sistem pengenalan sidik jari adalah sebagai berikut:

1. Membuat referensi sidik jari yang lebih banyak sehingga hasil yang diperoleh menjadi lebih akurat.
2. Untuk pengembangan selanjutnya pengenalan sidik jari dapat dilakukan secara *real time*.

#### PUSTAKA

- [1] A. Howard, *Dasar-Dasar Aljabar Linear*. Drexel University.
- [2] A tutorial on Principal Components Analysis, [http://csnet.otago.ac.nz/cosc453/student\\_tutorials/principal\\_components.pdf](http://csnet.otago.ac.nz/cosc453/student_tutorials/principal_components.pdf), 21 September 2006.
- [3] Fingerprint, <http://en.wikipedia.org/wiki/Fingerprint>, 6 Januari 2007.
- [4] Fingerprint Classification, [http://www.biologycorner.com/worksheets/fingerprint\\_class.html](http://www.biologycorner.com/worksheets/fingerprint_class.html), 21 Agustus 2006.
- [5] Fingerprint Patterns, [http://www.reachoutmichigan.org/funexperiments/agesubject/lessons/handouts/print\\_patterns.html](http://www.reachoutmichigan.org/funexperiments/agesubject/lessons/handouts/print_patterns.html), 21 Agustus 2006.
- [6] Hamming distance, [en.wikipedia.org/wiki/Hamming\\_distance](http://en.wikipedia.org/wiki/Hamming_distance), 15 November 2006
- [7] Integrated Automated Fingerprint Identification System, <http://en.wikipedia.org/wiki/IAFIS>, 6 Januari 2007.
- [8] J. Hollmen, Principal component analysis, <http://www.cis.hut.fi/jhollmen/dippa/node30.html>, 12 September 2006.
- [9] M Revilla, The Hamming Distance Problem, [acm.uva.es/problemset/v7/729.html](http://acm.uva.es/problemset/v7/729.html), 16 November 2006.
- [10] Principal components analysis, [en.wikipedia.org/wiki/Principal\\_components\\_analysis](http://en.wikipedia.org/wiki/Principal_components_analysis), 12 September 2006.
- [11] S. Bhatti, Channel coding; Hamming distance, [www.cs.ucl.ac.uk/staff/S.Bhatti/D51-notes/node30.html](http://www.cs.ucl.ac.uk/staff/S.Bhatti/D51-notes/node30.html), 15 November 2006.